NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

EVALUATION OF THE IMPACT OF MULTISPECTRAL IMAGE FUSION ON HUMAN PERFORMANCE IN GLOBAL SCENE PROCESSING

by

Brice Landreau White

March, 1998

Thesis Advisor: William K. Krebs
Second Reader: Harold J. Larson

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1998	Master's Thesis

4. TITLE AND SUBTITLE
EVALUATION OF THE IMPACT OF MULTISPECTRAL IMAGE FUSION ON HUMAN PERFORMANCE IN GLOBAL SCENE PROCESSING

5. FUNDING NUMBERS
ONR#N0001497WR30078
ONR#N0001497WR30091

6. AUTHOR(S)
White, Brice Landreau

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research, Arlington, VA
Lockheed Martin E&M, Orlando, FL

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*. An observer extracts local and global information from a natural scene to form a visual perception. Neisser (1967) and Treisman (1985) demonstrated that a natural scene contains different types of features, i.e., color, edges, luminance, and orientation to aid visual search. Infrared and visible sensors present nighttime images to an observer to aid target detection. These sensors present the observer an adequate representation of a nighttime scene, but sometimes fail to provide quality features for accurate visual perception. The purpose of this thesis is to investigate whether color features (combining an infrared and visible sensor image) improve visual scene comprehension compared to single-band grayscale features during a signal detection task. Twenty-three scenes were briefly presented in four different sensor formats (infrared, visible, fused monochrome, and fused color) to measure subjects' global visual ability to detect whether a natural scene was right side up or upside down. Subjects are significantly more accurate at detecting scene orientation for an infrared and fused color scene compared to a fused monochrome and visible scene. Both the infrared and fused color sensor formats provide enough essential features to allow an observer to perceptually organize a complex nighttime scene.

14. SUBJECT TERMS	15. NUMBER OF PAGES
IMAGE FUSION, PREATTENTIVE PROCESSING, VISUAL PERCEPTION	66
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified/A	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

EVALUATION OF THE IMPACT OF MULTISPECTRAL IMAGE FUSION ON HUMAN PERFORMANCE IN GLOBAL SCENE PROCESSING

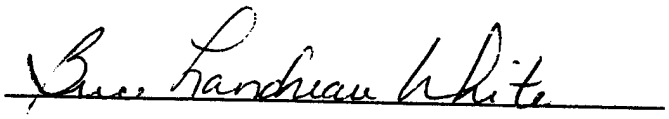
Brice Landreau White
Lieutenant Commander, United States Navy
B.S., Virginia Military Institute, 1984

Submitted in partial fulfillment of the
requirements for the degree of

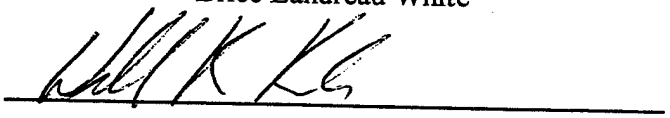
MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

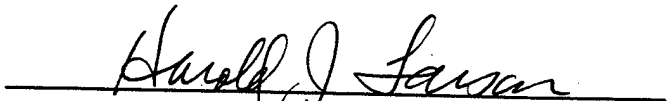
NAVAL POSTGRADUATE SCHOOL
March 1998

Author: 

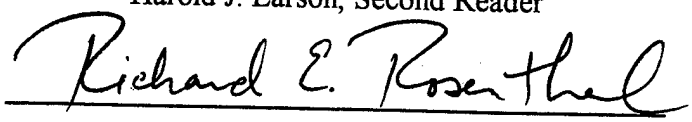
Brice Landreau White

Approved by: 

William K. Krebs, Thesis Advisor



Harold J. Larson, Second Reader



Richard E. Rosenthal, Chair
Department of Operations Research

ABSTRACT

An observer extracts local and global information from a natural scene to form a visual perception. Neisser (1967) and Treisman (1985) demonstrated that a natural scene contains different types of features, i.e., color, edges, luminance, and orientation to aid visual search. Infrared and visible sensors present nighttime images to an observer to aid target detection. These sensors present the observer an adequate representation of a nighttime scene, but sometimes fail to provide quality features for accurate visual perception. The purpose of this thesis is to investigate whether color features (combining an infrared and visible sensor image) improve visual scene comprehension compared to single-band grayscale features during a signal detection task. Twenty-three scenes were briefly presented in four different sensor formats (infrared, visible, fused monochrome, and fused color) to measure subjects' global visual ability to detect whether a natural scene was right side up or upside down. Subjects are significantly more accurate at detecting scene orientation for an infrared and fused color scene compared to a fused monochrome and visible scene. Both the infrared and fused color sensor formats provide enough essential features to allow an observer to perceptually organize a complex nighttime scene.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

For decades, military forces have used visual detection devices to enhance the ability to detect potential targets. The military operator has used long-wave infrared energy for target detection and visible to near-infrared energy for situational awareness. Each sensor provides the operator with important characteristics of the scene; however, in some cases, ambiguous information between the two sensor outputs may inhibit operators' detection sensitivity. In searching for a target, the operator must choose between the two sensor outputs to determine which sensor displays the target. Due to the spectral difference, the two sensors may display the scene in a completely different way. The operator must determine if the object displayed by each sensor is a target or noise. The inconsistent information between the two displays makes this task difficult, and thus the operator must guess whether the object is a target or noise. To reduce operator confusion, researchers have postulated that combining two spectral sensors into a single fused scene will improve operator detection sensitivity (Palmer, Ryan, and Tinkler, and Creswick, 1993). Palmer et al. (1993) demonstrated that pilots' situational awareness may benefit from a fused first generation long-wave infrared and an image-intensified charged coupled device displayed on a UH-1 aircraft.

The majority of image fusion algorithms attempt to improve operator detection by increasing target contrast (Scribner, Satyshur, and Kruer, 1993; Ryan and Tinkler, 1995; Therrien, Scrofani, and Krebs, 1997; Waxman, Gove, Fay, Racamato, Carrick, Seibert, and Savoye, 1996a). The fundamental objective for each of these fusion algorithms is to extract, pixel-by-pixel, the most important information from each spectral sensor, enhance each pixel element, and then combine the two elements into an enhanced scene. The algorithms use contrast between pixels to delineate features imbedded in the scene, but they differ in the method used to enhance the pixel elements. Each algorithm uses a different monochrome gray scale or artificial color method to enhance the pixel information.

Researchers have used visual search experiments in an effort to show that image fusion improves target detection time and accuracy. These visual search experiments have provided inconsistent results, regardless of the fusion algorithm used, due to the interaction between the spectral sensors and scene characteristics. Inconsistencies are reflected by both the time required to detect a target and the accuracy with which that target is detected. In a guided visual search task, the ability to quickly and accurately detect a target depends on the context of the entire scene. The more coherent and logical the scene context, the more accurately a subject can understand the scene; the result is an improved visual search task (Boyce, Pollatsek 1992a, Wolfe 1994). A more fundamental research approach would be to investigate the improvement in global scene perception due to image fusion versus improvements in target detection. If improvements in global scene perception due to image fusion can be shown to be consistent from scene to scene, then that consistent behavior can be translated to the commencement of the visual search task. Additionally, the insight gained from investigating global perception could be used to refine fusion algorithms and improve scene to scene consistency in target search and detection. The purpose of this research is to investigate the effect of image fusion on the processing of the global scene context.

The experiment was conducted on a Pentium 200-megahertz computer using Vision Works 3.0.4 (VRG) graphics display software. The video system consisted of a Cambridge VSG 2/4 controller video board and a FlexScan FX-E7 21-inch video monitor. The high resolution FlexScan FX-E7 color monitor had a 21-by-20-inch display area and an anti-reflective, non-glare, P-22 phoshere CRT. The CRT resolution was 600 by 800 pixels, with 75.02 x and 74.92 y pixel per degree. The CRT operated at 98.9 msec frame update rate and used an 8-bit look-up table (LUT) to control the red, blue and green guns. The CRT was positioned 1.0 meter from the subject, with the viewing distance and angle maintained by an adjustable chin rest.

Source images were obtained from the evaluation test flights of the Texas Instruments Image Fusion System (IFS) for the Army Night Vision and Electronic Sensors Directorate (NVSED). Intensified (I^2) and infrared (IR) images were selected from

available video footage, and still images were obtained during the test flights. The images were chosen to fit into five categories: man-made, wooded, roads, sea and general. Each image contained a discernible horizon, which was located within one-fourth of the center of the scene. The Naval Postgraduate School (NPS) monochrome fusion algorithm, used to produce a third set of images, further manipulated the component I^2 and IR images. Fusing the original images and coloring them using the Naval Research Laboratory (NRL) algorithm created the final set of images. The images were cropped to the size of the smallest image available. Size consistency ensured that no scene provided more information than that which was based solely on image size. The images were then inverted to produce the upside-down version of the image for the experiment.

The 920 trial images were composed of four independent variables. Each subject viewed both the right-side-up and the upside-down image, providing 460 trials of each type. For both image orientations, all four sensors were represented. The resulting 115 image trials were shown as 23 scenes over five Stimulus Onset Asynchrony (SOA) conditions. The 23 scenes were then broken into two sessions, the first one containing 12 scenes and the second session 11 scenes. For example, in the first session, 12 scenes were shown in each of the four sensor formats and both orientations to produce a block of 96 trials. Each block was then replicated five times for each condition of SOA. The images were presented in random order in each block during a given session. With sixteen subjects, each viewing 920 trial images, a total of 14,720 trials were conducted.

Analysis of Variance showed that the accuracy with which a subject can determine orientation of a scene is highly dependent on the sensor used and context of the scene. The interaction from scene to scene is driven primarily by the complexity of the scene. The more complex the scene, the more accurately the subject perceived the orientation. Image complexity consists of the scene context (e.g., levels of color, shape, curvature, object orientation) and is not related to the number of elements present in the scene. Additionally, the ability of a subject to accurately determine orientation of a scene was statistically dependent on prior experience, but not practically significant. Prior proficiency did not give subjects a distinct advantage in determining the overall scene

context. This suggests that scene perception is processed preattentively and is not learned or improved by training. If scene perception is indeed independent of proficiency, then image fusion provides a basic advantage for the commencement of guided search. An additional value of this study is the method it provides for further investigation of a scene context effect for search and detection of targets imbedded in scenes.

ACKNOWLEDGMENT

First and foremost, the author would like to acknowledge the Lord, Jesus Christ for His unending inspiration, guidance and strength to finish this endeavor.

Completion of this thesis is due to the unwavering support and sacrifice made by my wife Dawn. A special thanks to Sarah Grace, Rachael, Nathanael, Josiah and Rebekah for their understanding and patience in always waiting for me.

A special thanks to Captain Petho for his insight and comments that helped in the completion of this thesis.

Finally, my appreciation to Professors William K. Krebs and Harold J. Larson for their extensive guidance and time commitments.

I. INTRODUCTION

The ability to detect and classify contacts is paramount in the safe execution of day-to-day submarine operations. In the deep underwater environment, a submarine uses sonar as the primary sensor to detect contacts. At periscope depth, a specialized depth zone, the sonar's ability to detect contacts is degraded. To compensate for the submarine's vulnerability, the crew relies on visual input from the periscope. The submariner can observe surface traffic with moderate illumination within the general area of the submarine; however the submarine must operate at a reduced speed, which limits maneuverability. As a result, the crew wastes time maneuvering around surface vessels, rather than focusing on potential surface or air hazards. A system that could detect surface and air targets at greater distances would enhance the crew's situational awareness and reduce the probability of a collision.

At submarine periscope depth, two key factors, field-of-view and sensor type, limit an operator's detection of an object. Typically, the conventional submarine periscope uses a visible sensor and an image-intensified camera, each possessing a narrow field-of-view. These two sensors are adequate during high illumination, but are severely degraded during low-light operations. Nighttime operations can be especially hazardous due to the limited visibility for navigating around obstacles located near a coastline. A well-trained, experienced watch section can overcome some of these limitations. However, without enough sensory input, the operator may miss something; e.g., a small vessel masked by shore lighting. If the sensor's signal-to-noise ratio is poor, the operator should not be blamed for the error. For example, during twilight operations in 1995, a submarine crew failed to detect a small surface contact off the coast of California. The crew lost spatial awareness and collided with the surface contact due to the poor characteristics of the visible sensor. Both vessels returned safely to port; however, the collision resulted in a significant breach to the surface vessel's hull and serious damage to the submarine's antennae. The submarine crew was held responsible, but this mishap might have been avoided if the crew had access to a better detection device.

Submarine manufacturers have developed a new periscope at the request of the Chief of Naval Operations Submarine Warfare division. This new periscope, called the "Photonics mast" incorporates a high-resolution infrared sensor and a visible sensor. Because of its poor signal-to-noise ratio, the image-intensified sensor was replaced with an improved infrared sensor. The infrared and visible spectral bands provide complementary information that helps the user operate effectively during both day and night operations. The operator will obtain important scene characteristics from each sensor and, based on a comparison between the two outputs, decide whether to reject or accept the object. The operator may have to make a quick decision between conflicting information provided by two sensors, and the right decision may not be evident. To increase the probability that the operator will make the correct response, the two sensor formats must be displayed with minimal ambiguity. Krebs, Scribner, Miller, Ogawa, and Shuler (1998) showed that combining an infrared sensor and an image-intensified camera into a single fused image improved aviators' target recognition. This integration of information is similar to that of pit vipers, which fuse visible and infrared information to increase situational awareness of their surrounding environment and, therefore, better detection of their prey (Newman and Hartline, 1982). It is hypothesized that a fused scene, compared to either the visible or infrared sensors alone, will improve the submariner's situational awareness during surface surveillance. If the operator could easily detect and classify surface vessels, then the submarine's crew would reduce the risk of a collision. Furthermore, the fused image may provide additional navigation information that a SEAL platoon could use for coastline insertion.

A. SENSORS

The visual and non-visual electromagnetic spectrum of interest spans from 400 nano meters (nm) to approximately 14 microns (μm). The normal human visible range is from 400 nm to 700 nm. The exploitation of the electromagnetic spectrum outside the visible range can enhance military operations considerably. One particular period during

which the ability to use the visible electromagnetic spectrum is degraded is after sunset, when the available light in the sky drops. The necessity to see "into the night" is the driving force behind the development of Night Vision Devices (NVDs). NVDs consist of two basic types: Image intensifier (I^2) and Infrared (IR). A brief overview of the relevant electromagnetic spectrum is presented to provide the basic knowledge necessary to understand the goal of this research. Marine Aviation Weapons and Tactics Squardron One (MAWTS-1, 1994) provides more detailed coverage of the theory of the electromagnetic spectrum.

1. Image Intensifiers

a. Sources

The visual portion of the electromagnetic spectrum, or optical radiation, is exploited by I^2 sensors which enables an operator to view images during low-light conditions. "Light, or optical radiation, manifests itself in two ways; as particles of energy called photons or as waves propagated through a medium, which in this case is air" (MAWTS-1, 1994). Sources of light are divided into two categories, generated and reflected. Luminance (reflected light), which originates from either terrain or man-made objects, is expressed in terms of foot-lamberts (ft-l). Illuminance (generated light), from a source such as the moon, is measured in lumens per square meter (lm/m^2) and expressed in terms of lux.

Several variables affect illumination magnitude and the strength of the resulting luminance. The amount of illumination available in the night sky is directly related to the phase, angle, and properties of the moon's surface. Stars and other celestial bodies that contribute to the light intensity of the night sky have significantly less effect than the moon. The sun contributes short-duration light to the night sky during the periods just before sunrise and just after sunset.

b. Degrading Factors

Atmospheric conditions have the greatest range of effect on the I^2 scene. Many atmospheric factors can significantly reduce the effectiveness of the I^2 scene, the dominant one being water vapor. The type and magnitude of water vapor present can be either invisible or opaque on an I^2 display. Such wide and rapid fluctuations in the effectiveness of I^2 due solely to water vapor require operators to have significant experience to recognize a deteriorating condition. To a novice operator, and even to an experienced operator, severe degradation in visibility can occur without the operator realizing it. Figure 1.1 depicts a scene from an external source, such as the moon, where a pilot will perceive photons reflected from the source.

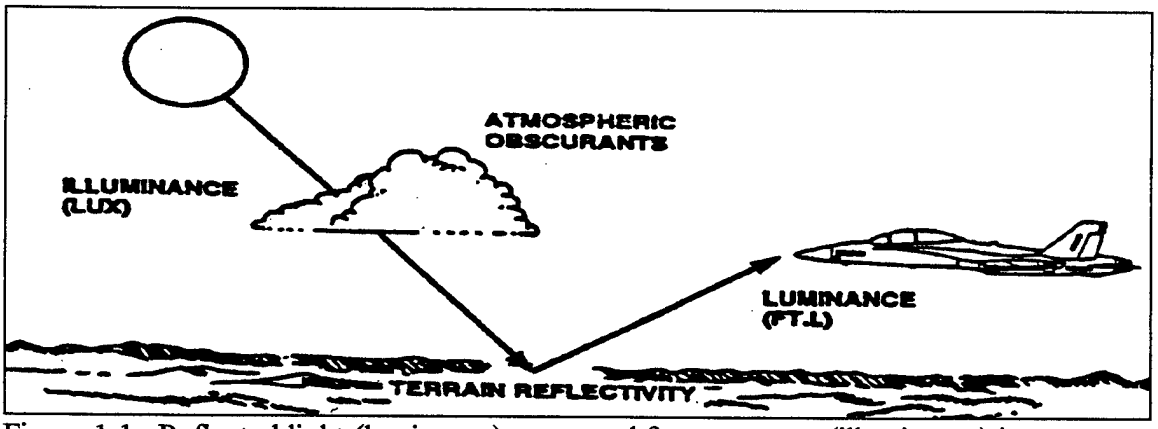


Figure 1.1. Reflected light (luminance) generated from a source (illuminance) is attenuated by atmospheric conditions and received and processed by a I^2 detector (MAWTS-1, 1994).

2. Infrared Sensors

a. Sources

All objects with a temperature above absolute zero (-273 degrees Celsius) emit energy, most of which is in the IR portion of the EM spectrum. An increase in temperature will increase an object's molecular vibrational motion, thereby increasing its energy state.

> When the elevated energy state collapses, thermal energy in the form of radiation is emitted (MAWTS-1, 1994).

Sources of an IR signature are of two types, man-made objects and stored solar energy. Man-made objects radiate an IR signature due to heat generated by friction, combustion of some type, and living creatures. Stored solar energy is radiated when the surrounding terrain or air cools. Unlike the visual spectrum, IR is independent of illumination and lumination. The magnitude of the IR signature from an object is controlled by two factors: conditions of the surrounding environment and thermal properties of the object. These two controlling factors can cause radical changes in the thermal picture hour to hour and make it difficult to quantitatively state how easy or difficult it will be to detect any one type of target.

b. Degrading Factors

Two primary factors can greatly reduce the effectiveness of the IR scene. The first, atmospheric conditions, has the widest range of effect on the IR scene due to energy absorption and scatter. The dominant factor impacting thermal signature loss is the absorption of thermal energy not scatter effects. Thermal absorption in the atmosphere is caused by a variety of gases and particulate matter. The dominant component affecting thermal absorption is water. Unlike the visual spectrum, IR is dependent mainly on the magnitude of water vapor present (humidity). As humidity increases, the environment becomes impenetrable to IR due to complete absorption of all thermal energy being radiated by an object and the associated background.

The next two factors affecting the clarity of an IR scene are the proximity and thermal strength of surrounding thermal signatures. Confounding thermal sources have some effect on the detectability of the target. The strength of the confounding sources is dependent upon the conditions of the surrounding environment and the thermal properties of the target. In contrast to the I^2 scene, the IR scene degradation is easily recognized by the pilot. However, pilot experience does play a significant role in the search and detection process for targets due to the wide quantitative changes for any given

type of target. As Figure 1.2 shows, when the thermal signature is radiated from a scene through the atmosphere, the IR sensor will observe reduced radiant energy of the target.

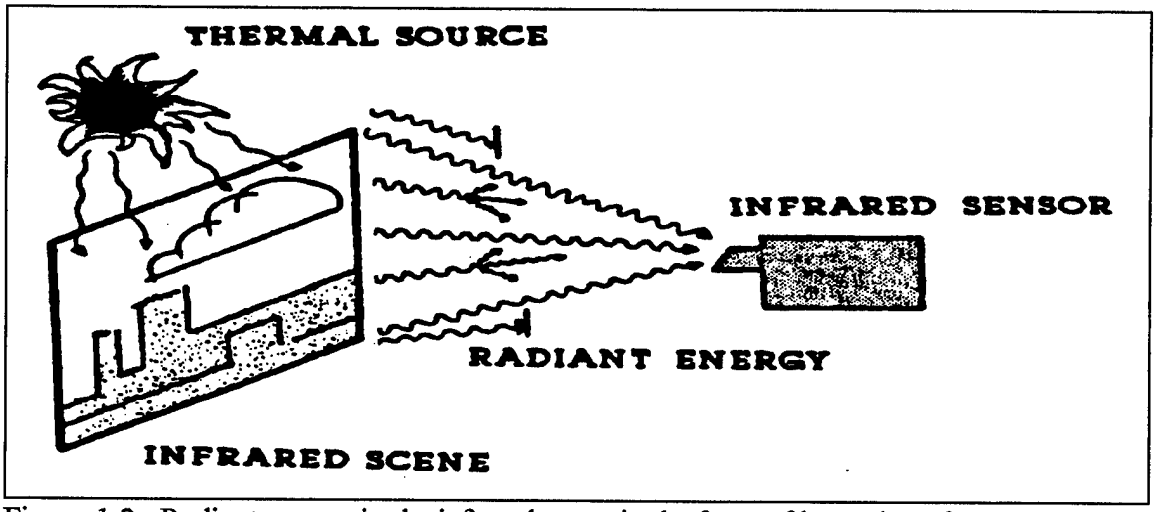


Figure 1.2. Radiant energy in the infrared scene in the form of latent heat from a source is radiated and attenuated in the atmosphere until detected and processed by a IR sensor (MAWTS-1, 1994).

B. SENSOR FUSION

1. Theory

The idea of fusing information from several sources (sensors) to obtain a global and more complete view of a scene has its basis in nature. The human nervous system operates five distinctly different senses. These individual sensors provide mutually exclusive information to the brain. The brain then integrates the input from the senses and creates a global view of the surrounding environment. Take, for example, the presence and proximity of a fire. In the case of fire, four of the five senses contribute information and help the individual determine the magnitude of danger and potential courses of action. When one of the four senses is impaired, say, vision obscured due to a closed door, the human body relies on one or more of the other senses to compensate for the visual impairment. The sensation of intense heat or dense smoke, or sound of a fire in close proximity, may provide sufficient information to allow safe egress from the dangerous

situation. A secondary solution is to attempt to improve the impaired sense — look under the door or even open it — both of which can have catastrophic results.

How the human body fuses information from the different senses to create a global representation of the surrounding environment can be used as a analogy to enhance the data obtained from available IR and I^2 sensors currently used for visual search. Continuing efforts to improve the individual sensors have provided substantially longer detection and identification ranges, resulting in an improved safety margin. The limitations of the electromagnetic spectrum cannot be avoided if we rely exclusively on the improvements of each sensor. Additional improvements may be obtained by extracting the advantages of each spectrum and fusing the sensor output into a single display. Both IR and I^2 degrade under different conditions; therefore, after the two are fused together, each should compensate for the other's shortcomings. The concept of fusing multiple electromagnetic wavelengths is inconsistent with human vision because the human eye is only sensitive to the visual spectrum. A better example is, again, the pit viper.

The pit viper has two unique senses to detect electromagnetic energy in the visible and IR spectrums. Processing this information allows the pit viper to hunt effectively in daytime, as well as at night. The snake uses these two distinct electromagnetic spectral sensors to identify and locate potential prey. Although the pit viper can strike with deadly accuracy on just the thermal signature of a potential prey, mapping of neurological impulses in its brain shows that the visual and IR spectrums are "fused" to provide a complete scene (Newman , Harline 1982).

2. Literature Review

Current research on the potential advantage of a fused image over the component images is centered around four types of experiments: paired comparison, method of equally appearing intervals, reaction time, and mean detection accuracy. The paired comparison experimental procedures have provided an excellent start in investigation of sensor fusion. In paired comparison experiments, subjects are presented a sequence of

image pairs and are forced to identify the "better" image. Subjects view fused images along with the component images in randomized order. Krebs, Buttrey, Lewis, and Mckenzie (1997) used a paired comparison experiment to determine the preferred sensor for use in a nighttime environment. The measure used was preference in locating a navigational cue in the scene. The sensor types included I^2, IR, and three different fused image techniques over 25 different scenes. The still-frame images were presented in pairs to obtain subject preference to the sensor type. Subjects preferred color fusion to the other sensor formats but there was a strong sensor by scene interaction. Sensors performed differently depending on the scene characteristics present (e.g., context, texture, and color).

A more robust qualitative evaluation is the method of equally appearing intervals. The evaluation of the night pilotage system known as the Advanced Helicopter Pilotage System (AHPS) provides insight into the advantage of image fusion (Ryan and Tinkler, 1995). The evaluation of the AHPS was based on a series of test flights with the AHPS providing real-time fused and component images to the aircraft in flight. During each flight, the crew conducted a series of night pilotage maneuvers with various sensor configurations. Each flight assessment consisted of a pilot evaluation for each sensor configuration, copilot evaluation of pilot performance, and an independent evaluator's observations and comments. The assessment concluded that the component sensors were "complimentary to one another and detect independent and unique characteristics of the scene" (Ryan and Tinkler, 1995). The pilots and evaluators used a grading scale from one to three, and the result showed that "image fusion was preferred over the individual sensors over all conditions and environments tested." (Ryan and Tinkler, 1995). Although the method of equally appearing intervals is an improvement over a simple paired comparison, it is still suspect because it depends completely on subjective responses. Therefore, because subjective preferences can be unreliable, a more quantitative approach to determine the efficiency of image fusion is desirable. An advantage in the ability to

search and detect target objects must be quantified to support the contention that image fusion is better than a single band representation.

Measuring search accuracy and reaction time provides a better basis to highlight the clear advantage of image fusion. Steele and Perconti (1997) conducted a series of experiments to determine the benefits of integrating gray scale and synthetic color for natural scene imagery. The independent variables were scene type, sensor type and the dependent measures were reaction time and accuracy. AHPS images were used to develop a series of laboratory experiments on horizon perception, recognition, and identification tasks. Still frames were produced from I^2, IR, and three different fused image techniques for a total of 25 different scenes. In the first experiment, the subjects' task was to determine whether or not the perceived scene orientation was level. Each scene was presented for approximately ten seconds. In their second experiment, videotaped sequences were presented to the subject. The subjects' task was to identify whether a predetermined target was embedded within the videotape sequence. Subjects reaction time and accuracy was measured for each trial.

Steele and Perconti (1997) showed that fusing different imaging systems is beneficial. Their experiment showed promising results, however it failed to demonstrate that sensor fusion is consistently superior compared to the component images. Their results showed that subjects' reaction time and accuracy was influenced by different combinations of sensor by scene type. This sensor by scene interaction indicated subjects did not benefit from color fusion across all scenes. A specific example of this confound is that, for some sensor formats, accuracy and reaction time did not correlate (i.e., the shortest response time did not have the highest accuracy) even though the images were viewed for as long as ten seconds. Steele and Perconti (1997) summarized that color fusion is not the best sensor type for all conditions, because the fusion algorithm is influenced by uncontrollable illumination, scene content, and color look-up-table parameters. As an alternative, color fusion may benefit targeting applications due to the added scene contrast.

A more critical investigation would be one that determines if the reaction time to detect a target can be improved using a fused image over the component images. Waxman, Gove, Seibert, Fay, Carrick, Racamato, Savoye, Burke, Reich, McGonagle, and Craig (1996) conducted an experiment to investigate the benefits of color fusion for target identification. Stimuli consisted I^2, IR, and fused-color nighttime scenes. A square target of varying contrast was imbedded within several heterogeneous natural scenes. Subjects detected low-contrast targets more rapidly in the fused scene compared to the component scenes. Waxman et al., (1996) concluded that subjects identified the fused color image significantly faster due to the opponent color qualities of the image. However, subje were slow to respond to I^2 and IR targets due to the poor target to scene contrast. Although these results show promise, there may be a methodological design flaw with the addition of a square target within the natural scene. This square target may cause the subject to experience some perceptual cognitive interference due to the introduction of a foreign geometric shape within a natural scene.

Another measure of the potential advantage of image fusion is the accuracy with which a target can be detected in real-world scenes. Toet, Ijspeert, Waxman, and Aguilar (1997) conducted an experiment to investigate whether the increased amount of detail in a fused image can improve a subject's ability to perform a situational awareness search task. Imagery was collected using visible and IR sensors. The images were taken in close proximity to dawn, when low light contrast and little or no thermal difference in scene background exist. During image collection, a target that was significantly higher in IR contrast (a person) was present in the scene. The collected images were fused into four additional image formats consisting of color and gray scale. The experiment consisted of computer-presented images followed by a schematic representation of the scene. Subjects identified target location, and then error distance and missed images were calculated. The results showed that color fused images performed better than monochrome fused and component images. The use of a low-contrast environment has two distinct effects. First, a high-contrast target in a homogeneous scene reduces the effect of distracters in the

scene and enhances detection capability. Second, the experiment shows that image fusion offers an advantage, but is limited by scene background, target scenario, and fusion algorithm used. The second result provides additional support for sensor-to-scene interaction.

C. THEORY OF TARGET RECOGNITION

The five senses of the human body provide constant sources of information on the surrounding environment. After the stimulus interacts with the specific sensor organ, the nervous system transfers the information to the brain, which interprets the electrical impulses and provides the necessary response. The entire process is known as "cognition" and refers to the way in which the human body senses, transfers, interprets, and responds to stimuli. A more focused type of cognitive process is humans' visual ability. Niesser (1967) offers the most concise definition of visual cognition: "Visual cognition, then, deals with the processes by which a perceived, remembered, and thought-about world is brought into being from as unpromising a beginning as the retinal patterns." The predominate model used is known as the "schema hypothesis" (Wolfe, 1994).

1. Visual Search

The schema hypothesis modeled by Wolfe (1994) encompasses the work of Niesser (1967), Treisman (1985) and others. The model's foundation rests on Niesser's proposal that visual search is a two-stage process composed of two fundamental components. The first process, which occurs before subject-driven search, that is before subject focuses attention on the object, is known as preattentive processing. The preattentive stage is dominated by a large-scale parallel visual search pattern. During the preattentive stage, the eye obtains a global perception of color, contrast, and texture. This global perception engages a visual schema which provides the necessary stimulus for the second stage of the visual process. The second stage is a serial process that determines the more-detailed features of a visual scene and identifies objects in the scene. The combination of these two visual stages provides the stimulus the brain needs to detect and

recognize objects. Niesser postulated that the transition between these two forms of visual search depends on "visual persistence" or "iconic memory." Visual persistence allows the eye to retain an icon of the image for about 500 milliseconds, which enables the human to "see" an image even after the image has been extinguished.

In Wolfe's (1994) model, the first stage, the preattentive stage, is a parallel process visual search. In a scene where features have "preattentive just noticeable difference" (Wolfe, 1994) the process is driven primarily by the stimulus and is a scene activated, bottom-up, process that is independent of the subject. Wolfe's model uses building blocks known as "feature maps" that determine how the visual search task is completed. In the preattentive stage, features such as size, color, and orientation are activated in separate feature maps. The weighted sum of the feature maps combine into an overall "activation map" that directs the attention of the subject to features in the visual scene. In a scene that does not provide the necessary noticeable difference between the object and background in a specific feature map, the parallel search process is not sufficient. In such conditions, a top-down, subject-controlled serial search is necessary. As in the parallel-only preattentive visual search, the search which uses both parallel and serial search results in an activation map that is a weighted combination of all the feature maps. Previous research has been "consistent with the idea that early visual analysis results in separate maps for separate properties, and these maps pool their activity across locations" (Triesman, 1985). Figure 1.3 depicts Wolfe's model.

The activation map is a representation of the relative magnitude of the stimuli in each location in the image and is derived from the feature maps. When attention is to be focused, the topographical analogy is that the hills would be higher on the priority list for visual attention. Conversely, the valleys in the activation map would be lower on the list. Wolfe concludes that the "activation map makes it possible to guide attention based on information from more than one feature. This is important in the search for targets not defined by a single unique feature" (Wolfe, 1994). As attention is focused and each location is identified, the next-highest peak is the next to be identified. This process

continues until either the target is identified or all points of interest are identified as non-targets. It has been shown that the pooling of these maps allows "rapid access to information about the presence of a target" (Triesman, 1985). Preattentive processing is dominated by a parallel search of coarse feature categories. The dominating scene attributes that control the preattentive stage are, in the order of effect, object size , color, motion, and orientation (Triesman, 1990).

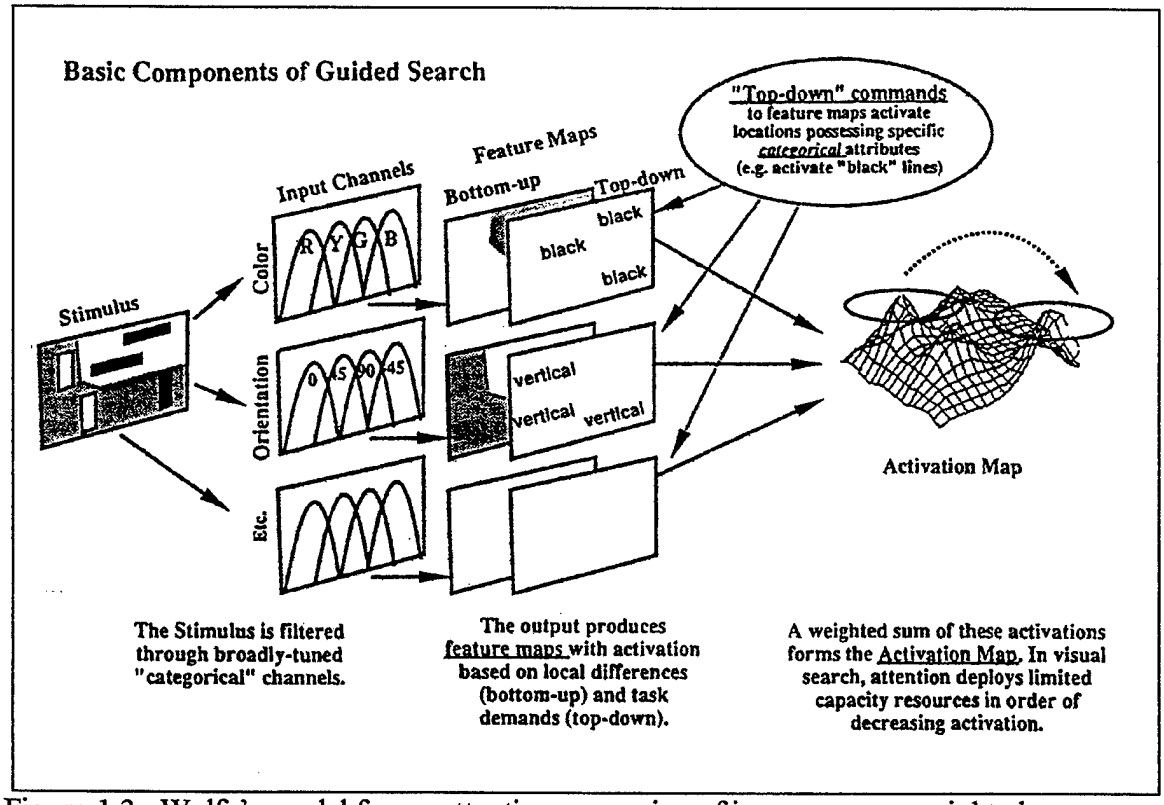


Figure 1.3. Wolfe's model for preattentive processing of images uses a weighted combination of feature maps to show the relative magnitude of the objects in the scene. The magnitude or "preattentive just noticeable difference" determines what features are detected preattentively and what features engage guided search (Wolfe 1994).

2. Visual Persistence

There is no doubt that a transition between parallel search and the subsequent serial search for features is built on volatile information storage. Subsequent research shows that the view that iconic memory and visual persistence are the same thing is incorrect (Phillips 1974, Coltheart 1980). Niesser stated that visual cognition is in itself a

"complex of processes" (Niesser, 1967). Niesser also alluded that, even when an image is terminated, "it seems apparent that the processing of iconic information can continue past this point." (Niesser, 1967). Refinement of Niesser's initial hypothesis shows that the overall process may be composed of both visual persistence and iconic memory (Coltheart, 1980).

The effects of stimulus luminance and duration on visual persistence provide the basis for differentiating visual persistence and iconic memory. Both Niesser's and Coltheart's theories depend on the assertion that the foundation that the persistence of an image, even after is has been extinguished, is due to residual neural activity. The length of time that the image persists depends on both the length of time the stimulus is exposed and the magnitude of the luminance. Further support for residual neural activity being the basis of visual persistence is that image complexity does not reliably affect the duration of the persistence (Irwin and Yeomans, 1991). Irwin and Yeomans concluded that "visible persistence appears to be a residual neural trace of an extinguished stimulus, rather than a byproduct of cognitively driven information-extraction process."

The duration of the visual persistence can last significantly longer than the stimulus itself. Duration of the image after the stimulus has been removed was initially determined to be approximately one second (Niesser, 1967). The duration of the image was later shown to exist for approximately 200 msec (Haber, 1970). In experimental testing, the control of visual persistence is vitally important when investigating the visual preattentive process. Masking is the tool that is readily available to provide the necessary control of the stimulus image persistence.

3. Backward Masking

Masking of a visual stimulus can be categorized in many different ways. The categories into which the masking falls depends on by the method used in the experiment. Masking can be performed in three time domains. The time domain is defined by the Stimulus Onset Asynchrony (SOA) or the time interval between stimulus offset and mask onset. The first time domain is "forward masking" or before stimulus onset. In forward

masking, the masking image is presented just before the test stimulus. Forward masking relies on the visual persistence of the masking image, allowing the mask image to "merge" with the test stimulus. In forward masking, the decay of the mask image after mask offset occurs requires that the proximity to the test stimulus be short. The second time domain is simultaneous masking, presenting the mask stimulus and test stimulus simultaneously. The final time domain is backward masking, when the mask stimulus is presented after the test stimulus. Figure 1.4 depicts backward masking.

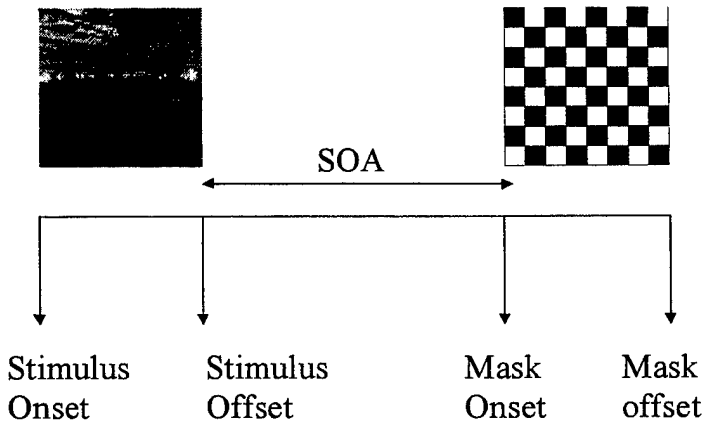


Figure 1.4. Backward masking of a visual scene with visual noise.

The three types of backward visual masking are masking by light, by visual noise, and by metacontrast. Kahneman (1968) provides a detailed review of all three types of backward masking. In general, the best method for controlling the persistence of an image is masking by noise. As Kahneman states, "When test stimulus is a form and masking stimulus is a field of visual noise, their joint presentations results in a general degrading of the image" (Kahneman, 1968). The effect of visual noise on the test stimulus is greatest when the masking stimulus is presented immediately following the test stimulus. Kahneman calls this "type-A." Additionally, in type-A masking, the effect of the masking stimulus can be demonstrated when the SOA is less than 100 msec (Niesser, 1967). Later

research showed that visual noise will have some effect until the SOA exceeds the visual persistence of the test stimulus, after which no effect can be seen. "What is most noteworthy is that there are no exceptions to the general principle that the visual noise mask will terminate the persistence of a stimulus if the mask arrives before the normal persistence is ended of its own accord" (Haber, 1970).

D. SCENE FACTORS

1. Context

Scene context is a key factor in the identification of a target in a natural scene. The scene context initiates a specific schema that directly impacts the speed at which targets in the scene can be named. The effect of scene context on object naming occurs early in scene viewing, typically on the first visual fixation (Boyce and Pollatsek, 1992). The context determined early in scene viewing is then used in later stages of visual search to identify the target. Many significant factors in the scene context affect the speed and accuracy with which the subject can identify a target object.

The general background characteristics of a visual scene are the driving factors behind a subject's global scene perception. In the preattentive stage, when coarse object characteristics are processed, the size and color of the background surrounding a target dominate the scene. Brief exposure experiments showed "that the background of a scene was the sole cause of the context effect" (Boyce and Pollatsek, 1992a). Previously, Biederman had shown that having a single inconsistent object in a visual scene did not affect the accuracy with which a target was detected (Biederman, 1982).

2. Scene Complexity

If the number of individual features of a specific object in a visual scene were increased, it would make sense to assume that the target would be more discernible than others in the scene. Early research proved that the "dimensions" or the feature maps that define an object do impact detection. Increasing the number of feature maps or

"dimensions" that define an object in a scene results in a lower search time and, therefore, an increase in the scene processing rate (Teichner and Mocharnuk, 1979). Wolfe provided additional evidence that the reaction time to find objects in a scene increased at different rates depending on the number and type of features present. He showed that as the number of features present in a visual scene was increased, the object identification time was significantly reduced. Additionally, the more dominant features of color by size combination provided better results than color by orientation, further supporting Triesman's proposed hierarchy (Wolfe, 1993).

3. Experience and Prior Expectations

If a subject's experience level with the type of stimulus were increased, the natural expectation would be that performance level in search and detection of the stimulus would also improve. In an effort to determine if experience plays a part in visual scene perception, Biederman, Teitelbaum, and Mezzanotte (1983) conducted a series of experiments in which the stimulus backgrounds for a target image changed during a experiment. The learning effect due to repetition continued with consistent values of magnitude and slope. Thus, Biederman et al. concluded:

> The implication of these results are that the mechanisms for perceiving and interpreting nondegraded real-world scenes are so quick and efficient that conditions can readily be found in which priming and prior exposures of substantial portions of the scenes are not helpful for perceiving and judging certain aspects of those scenes (Biederman et al 1983).

E. HYPOTHESIS

The benefits of color fusion have been inconsistent (Krebs et al. 1997, Steele and Perconti 1997, Waxman et al. 1996, Toet et al. 1997). Researchers have been unsuccessful in clearly demonstrating that observers' detection for a color-fused target will be significantly better than an I^2 or IR target. Both I^2 and IR sensors suffer from poor dynamic range and uncontrollable factors, such as environmental and terrain conditions. Alternatively, color fusion is dependent upon the quality of input imagery. If the I^2 and IR

sensor input is good (or poor) then fusion will not show a significant improvement. However, if the I^2 and IR sensor input is mixed, I^2 good and IR bad, then fusion should be superior compared to the component bands. Previous studies (Krebs et al. 1997, Steele and Perconti 1997, Waxman et al. 1996, Toet et al. 1997) failed to demonstrate the benefits of color fusion due methodological design errors. Specifically, these studies required subjects to detect or identify a specific target within a natural scene. This experimental method was susceptible to sensor by scene interaction. The color fusion target scores were collapsed across the different scene conditions, thus nullifying the benefits of a color target for a particular scene type. As a result, this study proposes to investigate an alternative experimental method to test the benefits of color fusion.

This alternative method is based on preattentive processing. It is hypothesized that a visual search task will be influenced by the qualities of the preattentive target. Previous studies failed to prevent subjects' guided search for scene features. As a result, these studies were measuring focal attention mechanisms rather than preattentive processing. In order to prevent guided search and test preattentive processing, the experimenter needs to control the stimulus characteristics. Image duration and the length of visual persistence can control guided search: if an image is presented briefly enough, the subjects' visual cognitive process is limited to the preattentive process; and if the image duration, when coupled with visual persistence, is sufficiently long to allow other cognitive processes to engage, then the single effect of preattentive processing is lost. The combination of target image duration and visual mask will contain the cognitive process to the preattentive stage.

When the cognitive process is controlled, the feature maps associated with an image and the effect image fusion has on the maps can be determined. The feature maps are dependent on the overall scene context. Different sensors present the associated features with varying clarity and accuracy. The sensor-to-scene effect of the intrinsic features must be adequately represented. Sufficient scenes must be used to ensure that the experiment looks across multiple real-world scene environments, and not just those that

have shown superior fusion performance. The ability to globally determine scene orientation is a measure of the preattentive process and the accuracy with which it occurs. By using multiple scenes presented both right side up and upside down, an experimenter can test subjects ability to preattentively process real-world scenes. If a subject can correctly identify the orientation of the scene more accurately with image fusion, then the fusion of component images is superior in the preattentive stage of visual cognition.

The experiment presented in the next chapter investigates the following two hypotheses: first, that there will be no difference in the mean accuracy of the various sensors; and second, that there will be no difference in mean accuracy of the various sensors based on the experience level of the subject. The goal of the experiment is to examine the plausibility of these hypotheses.

II. METHODS

A. SUBJECTS

Sixteen male U.S. military officers volunteered to be subjects. Subject age ranged from 27 to 45 with a mean age of 32 years. All subjects had normal or corrected-to-normal visual acuity (20/20) and were naive as to the purpose of the experiment. Ten subjects had completed flight training either as aviators or flight officers. Half of the subjects were experienced, familiar with both component sensors (I^2 and IR), and proficient in the use of at least one of the sensors. All subjects signed an informed consent and were briefed on the ethical conduct for subject participation specified in the Protection of Human Subjects, SECNAV Instruction 3900.39B.

B. EQUIPMENT

The experiment was conducted on a Pentium 200-megahertz computer using Vision Works 3.0.4 (VRG) graphics display software. The video system consisted of a Cambridge VSG 2/4 controller video board and a FlexScan FX-E7 21-inch video monitor. The high resolution FlexScan FX-E7 color monitor had a 21- by 20-inch display area and an anti-reflective, non-glare, P-22 phoshere CRT. The CRT resolution was 600 by 800 pixels with 75.02 x and 74.92 y pixel per degree. The CRT operated at 98.9 msec frame update rate and used an 8-bit look-up table (LUT) to control the red, blue and green guns. The CRT was positioned 1.0 meters from the subject, with the viewing distance and angle maintained by an adjustable chin rest. A small floor lamp (5.51 cd/m^2) was positioned on the floor behind the CRT to reduce screen glare.

C. STIMULI

Source images were obtained from the evaluation test flights of the Texas Instruments Image Fusion System (IFS) for the Army Night Vision and Electronic Sensors Directorate (NVSED). Images were selected from available video footage and still images obtained during the test flights. The images were chosen to fit into five categories: man-

made, wooded, roads, sea, and general. Each image contained a discernible horizon, which was located within one-fourth of the center of the scene.

The Naval Postgraduate School (NPS) monochrome fusion algorithm used a modified Peli-Lim algorithm (Therrien, Scrofani, and Krebs, 1997). The NPS fused algorithm separated the scene into low-pass (low luminance value pixels) and high-pass components (high luminance value pixels). The high-pass component was amplified by a multiplication of the local luminance mean. The low-pass component was passed through a nonlinear luminance transformation so that the combination of the high-pass component would not saturate the fused scene.

The Naval Research Laboratory (NRL) color fusion algorithm used a two-dimensional false color based on principle component analysis (Scribner, Satyshur, and Kruer, 1993). The IR and I^2 images were digitized and plotted on a two-dimensional plot. The distributions of intensity values from both bands have a spheroid distribution extending along the principal component distribution. The elongated axis, principal component direction, represents brightness and the orthogonal axis is the chromaticity plane. The orthogonal direction was transformed into an expanded polar coordinate to generate a color circle with hue and saturation (Krebs, Scribner, Miller, Ogawa, and Shuler, 1998, pp. 6-7). For these images, red and cyan were used as the opponent colors on the color circle. For example, the IR used a red coloration on the bright pixels, which denoted the hotter objects in the scene (white hot IR configuration), while I^2 used a cyan color on pixels with more luminance. The two images were then fused to create a two-dimensional, artificially colored image. The final image colors correlate as: red - a hot IR (red) and dark I^2, cyan - bright I^2 and cold IR (dark), white - hot IR (red) and bright I^2 (cyan) and black - cold IR (dark) and dark I^2.

The I^2, IR, NPS, and NRL images were cropped to a standard size of 495 by 443 pixels. Size consistency ensured that no scene provided more information based solely on image size. The images were then inverted to produce the upside-down version of the image for the experiment. Figure 2.1 illustrates four different sensor formats— sensor by

scene type. The checkerboard mask, 500 by 450 pixels, consisted of alternating black and white inner squares, 10 by 10 pixels, in size. The mask mean luminance was 7.64 cd/m^2.

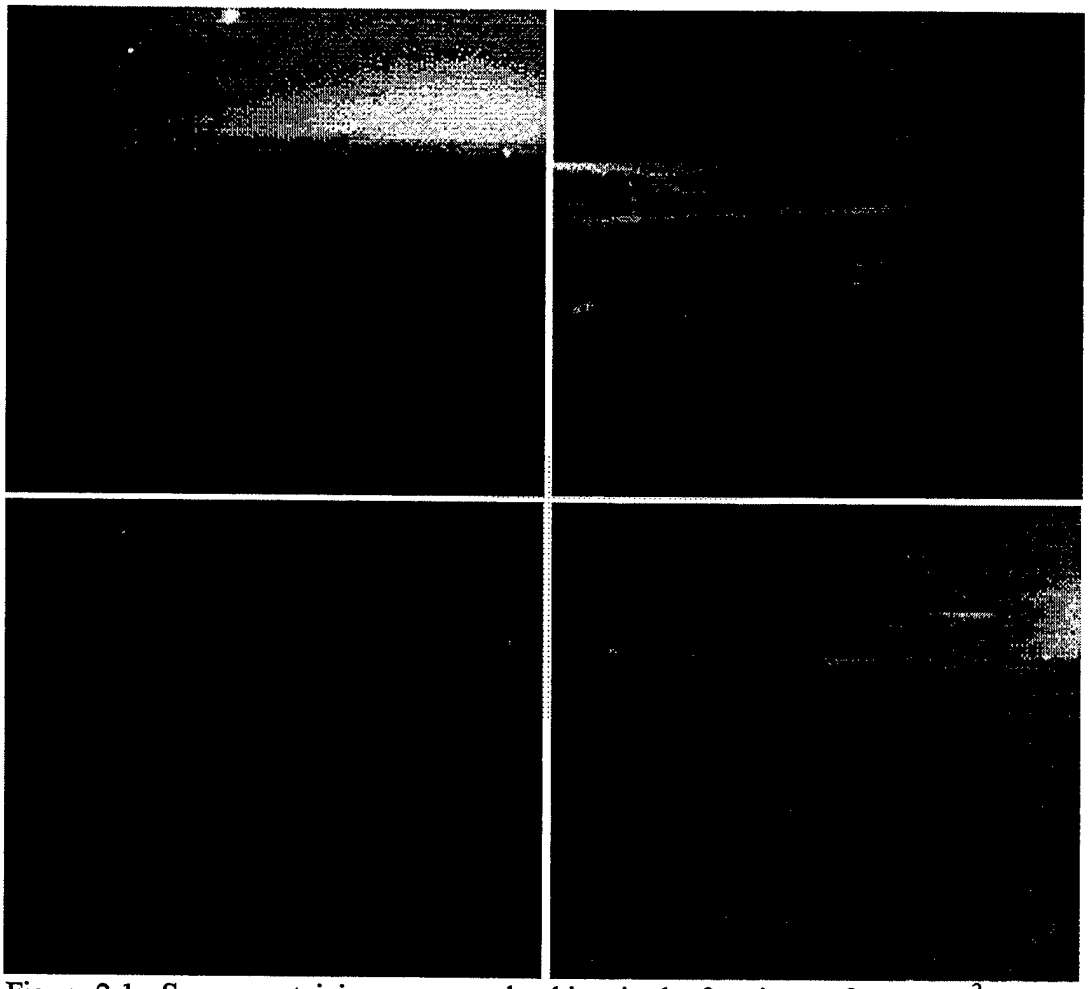


Figure 2.1. Scene containing a manmade object in the four image formats: I^2 (upper left), IR (upper right), NPS fused monochrome (lower left) and NRL fused color (lower right). Images courtesy of NVSED.

D. PROCEDURE

The experiment consisted of two sessions for each subject. Each session lasted approximately 45 minutes. At the beginning of each session, the subject received verbal instructions for the task, followed by forty practice trials. These practice trials were used to familiarize the subject with the task. The practice trials' stimuli were not used in the experimental trials.

The experimental trial consisted of the subject viewing a fixation cross located in the center of the screen. The subject initiated a trial by pressing either "1" or "2" on the keyboard. The fixation cross would extinguish, and after a 20 millisecond delay, the stimulus would be presented for 50 milliseconds. The stimulus was followed by a checkerboard mask after a predetermined SOA (0, 20, 40, 60, and 80 milliseconds). The checkerboard mask was presented for 50 milliseconds. At the completion of the stimulus sequence, the subject made a keyboard response as to the orientation of the scene. Subjects received aural feedback for incorrect responses. Figure 2.2 illustrates a single trial.

Each subject received 920 experimental trials composed of four different independent variables (orientation by sensor types, by scene types, by SOA). The 920 trials were divided into 460 right-side-up and 460 upside-down image formats. Each subject contributed one threshold point ("1" for correct and "0" for incorrect) for each orientation by four sensor types by twenty-three scene types by five SOA conditions across the two sessions. The twenty-three scenes were divided into two sessions, the first containing 12 scenes and the second containing 11 scenes. For example, in the first session, 12 scenes were shown in each of the four sensor formats by two orientations to produce a block of 96 trials for one SOA. The trials were randomly presented within each block by session across all subjects. In summary, each of the sixteen subjects viewed 920 experimental trials for a total of 14,720 trials.

The threshold points were grouped by subject experience level providing two databases (experienced and novice). Each data base cell contained one threshold point for each subject for a total of eight threshold points per cell and was categorized by orientation, sensor, scene and SOA. The threshold points were converted to a percentage which represented the accuracy that an individual, with a given experience level, determined scene orientation for each orientation, sensor, scene and SOA. These accuracy fractions or proportions are analyzed below.

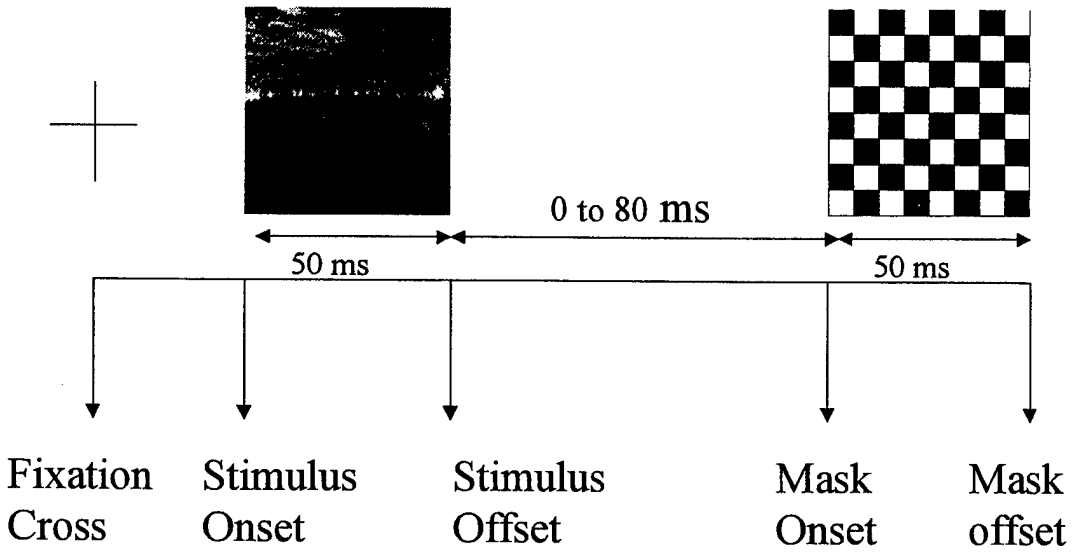


Figure 2.2. The experimental procedure. A fixation cross was presented on a blank screen for 50 msec. The stimulus followed and was presented for 50 msec. An SOA varying from 0 to 80 msec in 20 msec increments was followed by a checkerboard-masking pattern. The mask was presented for 50 msec and extinguished

III. RESULTS

A. DATA ANALYSIS

A completely within-subject design, Analysis of Variance (Sensor by Orientation by SOA by Scene by Experience) was carried out on the dependent measure accuracy. The results quoted are based on the raw percent accuracy contained in each data cell. The arc sine transformation, commonly used to stabilize variance, was also employed and gave essentially the same results. The large sample size available for analysis was expected to provide statistical significance even in cases when mean accuracy differed by only one or two percent. When conducting human response experiments, statistical significance based on such small differences in mean accuracy cannot be the sole factor used when determining significance of the independent variable effects or interactions. Therefore, practical significance must be considered when analyzing the data in this experiment.

There was a significant sensor main effect (F (3,1624) = 69.5567, p = 0.0000). Figure 3.1 illustrates that subjects responded more accurately to fused color and IR than to I^2 and FM. The boxplots show that the IR and fused color sensor types are similar in mean and interquartile range, and I^2 and fused monochrome have lower means and larger interquartile ranges. Tukey's method of multiple pair-wise comparisons showed that all sensor comparisons were significant (refer to Appendices A and B).

Analysis of the combined effect of the fused sensors was conducted following collation of the data into fused and non-fused sensor categories. There was not a significant sensor main effect when comparing both fused and both non-fused sensors, t (1838) = 0.4292, p = 0.6678. There was a significant sensor main effect when comparing only fused color and non-fused sensors, t (1378) = 3.754, p = 0.0008. In the case of fused color, the higher mean accuracy is expected due to fusion taking advantage of the better qualities of both component images (refer to Appendix B).

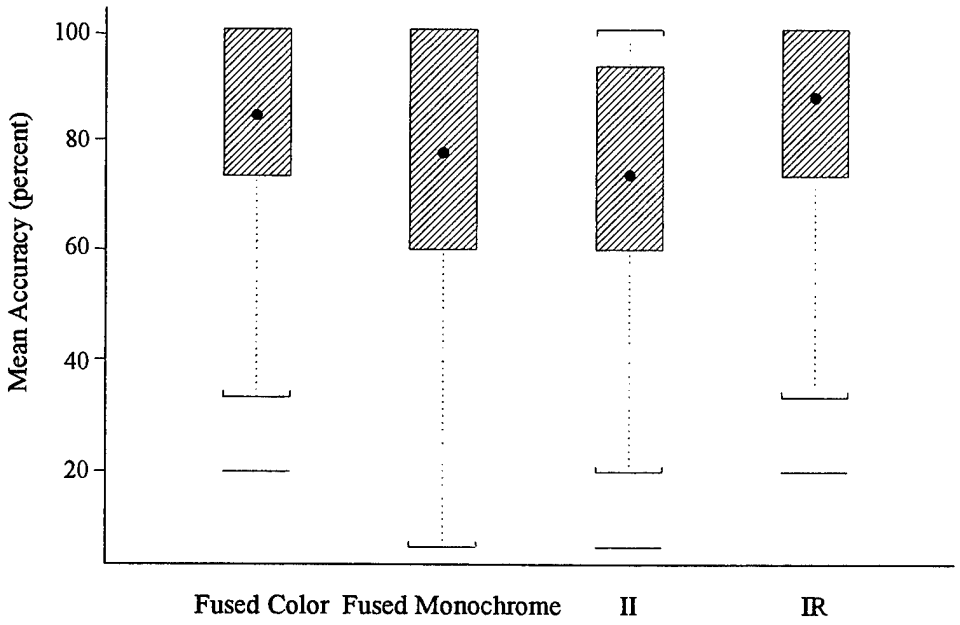


Figure 3.1. A significant main effect for sensor (F (3,184) = 110.0411, p = 0.0000) with accuracy as the dependent measure. The Box Plot shows the mean (dot) and interquartile range for sensor accuracy.

There was a significant experience main effect (F (1,1624) = 5.2869, p = 0.0216). Figure 3.2 illustrates that experienced subjects responded more accurately than inexperienced subjects. The boxplots show that the experienced and inexperienced subjects have similar means and interquartile ranges. The actual mean accuracy was 0.8285 and 0.8120 respectfully, which hardly differ in any practical sense.

There was a significant orientation main effect (F (1,1624) = 14.4939, p = 0.0001). Figure 3.3 illustrates that subjects responded more accurately to upside-down than to the right-side-up scenes. The boxplots show that the upside-down and right-side-up images have similar means and interquartile ranges. The actual mean accuracy was 0.8065 and 0.8340 respectfully, which hardly differ in any practical sense.

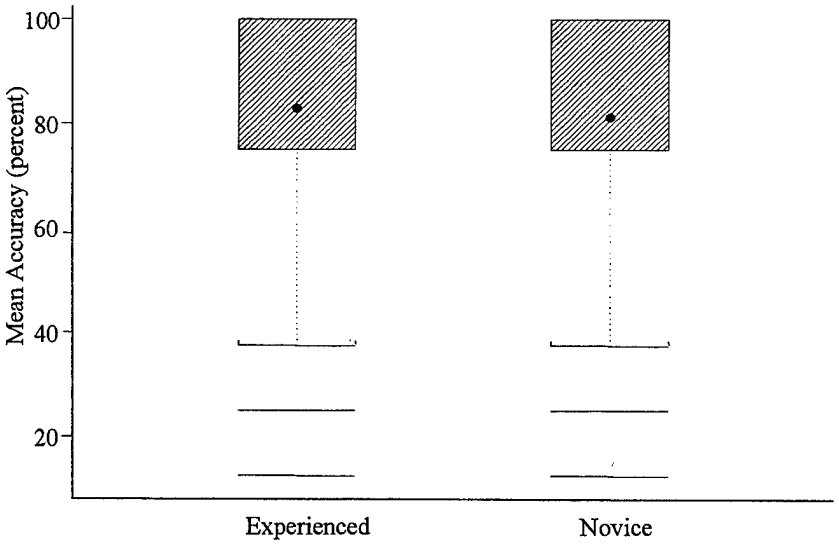


Figure 3.2. A significant main effect for experience ($\underline{F}$ (1,184) = 6.5387, $\underline{p}$ = 0.0114). The Box Plot shows the mean (dot) and interquartile range for experience accuracy.

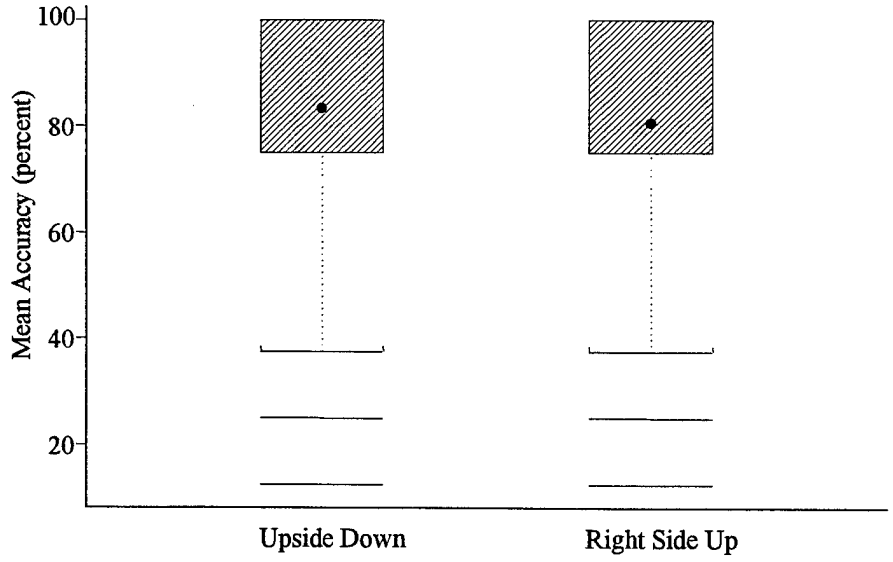


Figure 3.3. A significant main effect for orientation ($\underline{F}$ (1,184) = 12.2592, $\underline{p}$ = 0.0006). The Box Plot shows the mean (dot) and interquartile range with respect to orientation accuracy.

There was a significant scene main effect (F (1,1624) = 151.5301, p = 0.0000). Figure 3.4 illustrates that subjects respond more accurately to scenes containing a road and least accurately to a general scene. The boxplots show that the means are quite different but the interquartile ranges are similar (except for road). Tukey's method of multiple pair-wise comparisons showed all scene comparisons are significant except the tree vs. sea comparison (refer to Appendices A and B).

There was a significant SOA main effect (F (4,1624) = 4.1989, p = 0.0022). Figure 3.5 illustrates that subjects respond more accurately when SOA is 80 msec and least accurately when SOA is 0 msec. The boxplots show that the means are different with a consistent upward trend, and the interquartile ranges are similar for all SOAs. Tukey's method of multiple pair-wise comparisons showed only comparisons, 0 msec vs. 60 msec and 0 msec vs. 80 msec, were significant (refer to Appendices A and B).

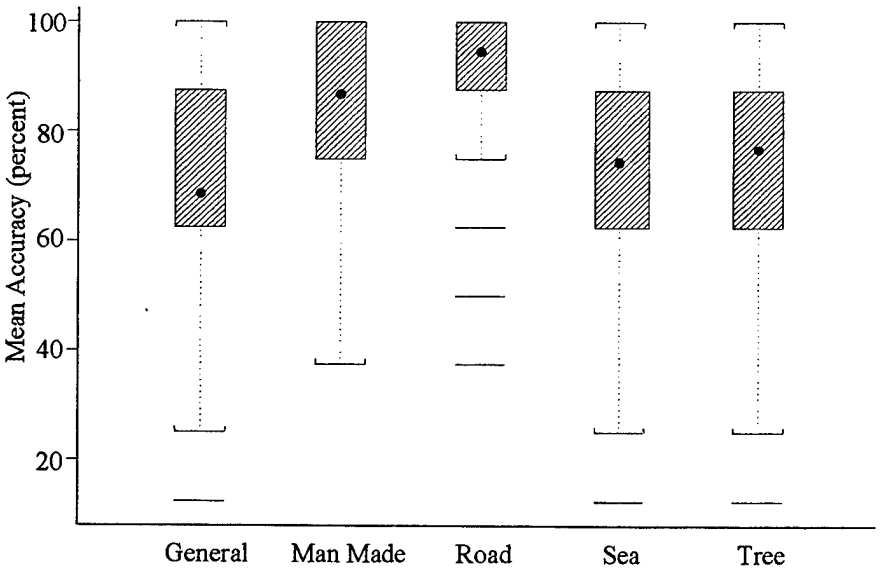


Figure 3.4. A significant main effect for scene category (F (1,1624) = 151.5301, p = 0.0000). The Box Plot shows the mean (dot) and interquartile range with respect to scene accuracy.

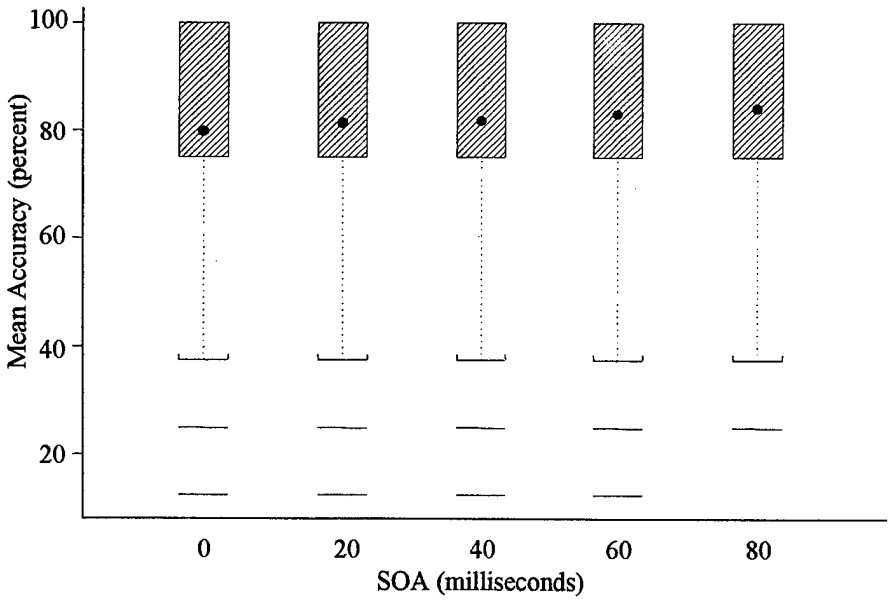


Figure 3.5. A significant main effect for SOA (F (4,1624) = 4.1989, p = 0.0022). The Box Plot shows the mean (dot) and interquartile range with respect to SOA accuracy.

There was a significant two-way interaction for scene by orientation (F (4,1624) = 7.7538, p = 0.0000). Figure 3.6 illustrates that different scene categories reacted in a consistent manner for both orientations with the exception of those categorized as tree scenes (refer to Appendix C). In the case of the tree category, the upside-down image was easier to detect than the right-side-up image.

There was a significant two-way interaction for sensor by orientation, (F (3,1624) = 5.0340, p = 0.0018). Figure 3.7 illustrates that an interaction exists for IR and FC by orientation (refer to Appendix C). In the case of IR and FC, the upside-down image is more accurately identified than the right-side-up image. Also, the IR and FC mean accuracy varies due to orientation and the fact that I^2 and FM are constant. There was a no significant interaction between sensor and experience (F (3,1624) = 0.7033, p = 0.5501). Figure 3.8 illustrates that no interaction exists for sensor by experience.

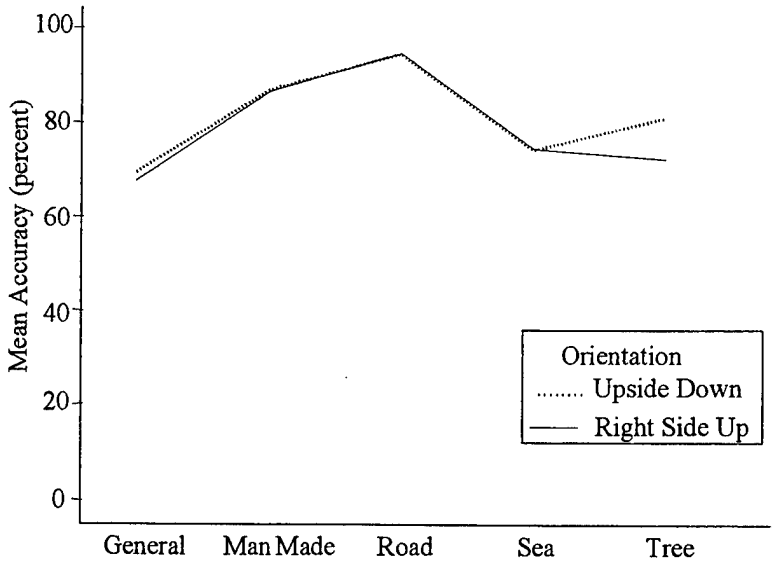


Figure 3.6. A significant interaction for scene by orientation (F (4,1624) = 7.7538, p = 0.0000).

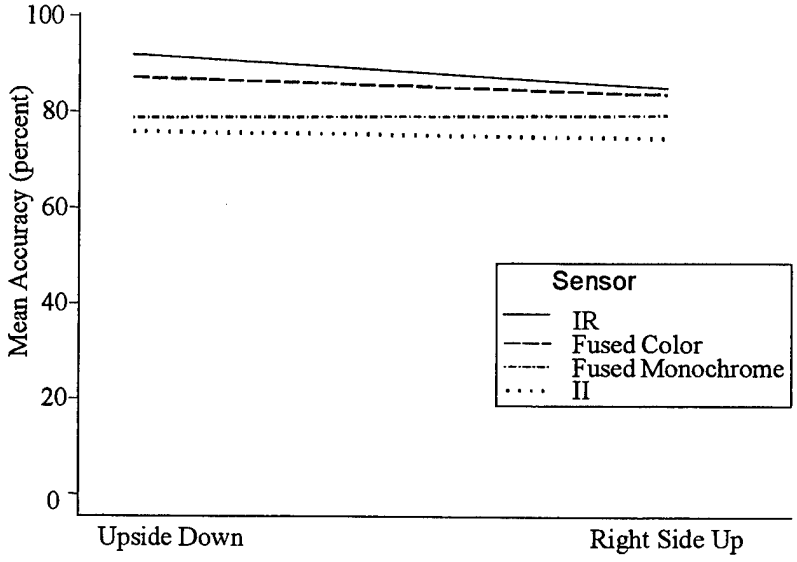


Figure 3.7. A significant interaction for sensor by orientation (F (3,1624) = 5.0340, p = 0.0018).

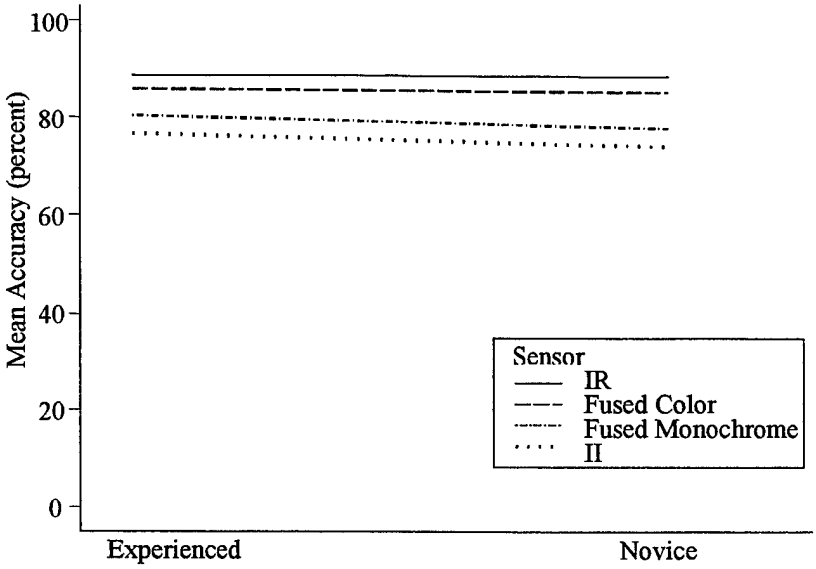


Figure 3.8. The experience by sensor interaction was not significant ($\underline{F}$ (3,1624) = 0.7033, $\underline{p}$ = 0.5501.

There was no significant interaction between sensor and SOA ($\underline{F}$ (12,1624) = 0.8922, $\underline{p}$ = 0.5545). Figure 3.9 illustrates that no interaction exists for SOA by sensor. There was a significant interaction between sensor and scene category ($\underline{F}$ (12,1624) = 4.9095, $\underline{p}$ = 0.0000). Figure 3.10 illustrates that IR and FC consistently performed better than FC and I^2 (refer to Appendix C). However, the magnitude of the improvement was not consistent.

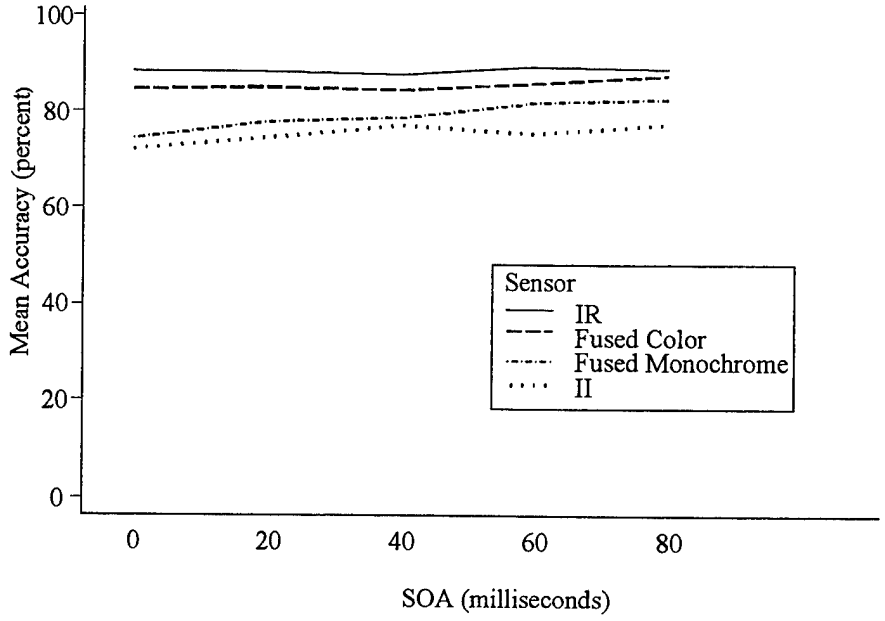


Figure 3.9. The SOA by sensor given interaction was not (F (12,1624) = 0.8922, p = 0.5545).

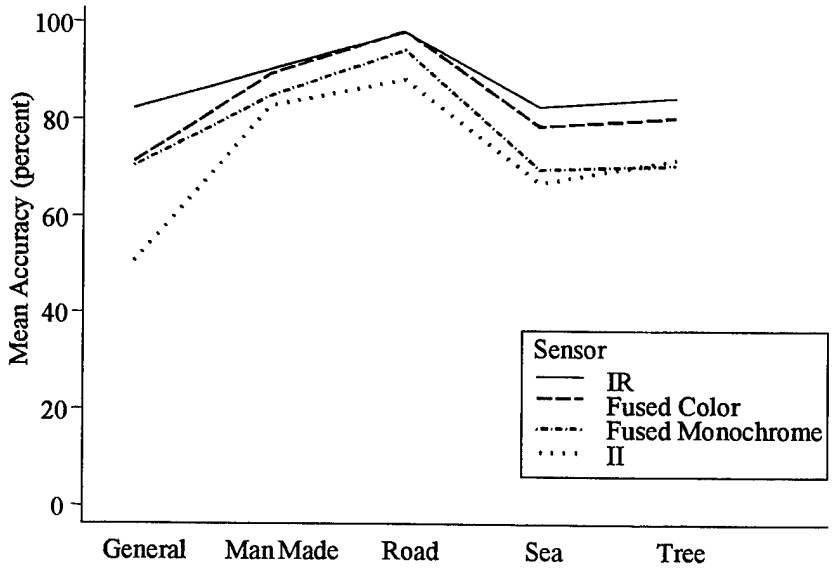


Figure 3.10. The sensor by scene interaction was significant (F (12,1624) = 4.9095, p = 0.0000).

B. POST HOC ANALYSIS

Post hoc analysis investigated the robustness of the experimental design. The experiment used 32 sessions, each consisting of five blocks. The blocks in any given session were randomized as to the SOA applied to the blocks (refer to Appendix D). Any significance of any of the five independent variables with mean accuracy would suggest a learning effect. The learning effect, if present, should produce a significant increase in the mean accuracy between the trial blocks.

A "within-trial" design showed a significant effect for the total mean accuracy by trial (F (4,14665) = 18.8296, p = 0.0000). Figure 3.11 illustrates that the statistical significance correlates to a 7.35-percent improvement in mean accuracy and does not support practical significance that learning occurred.

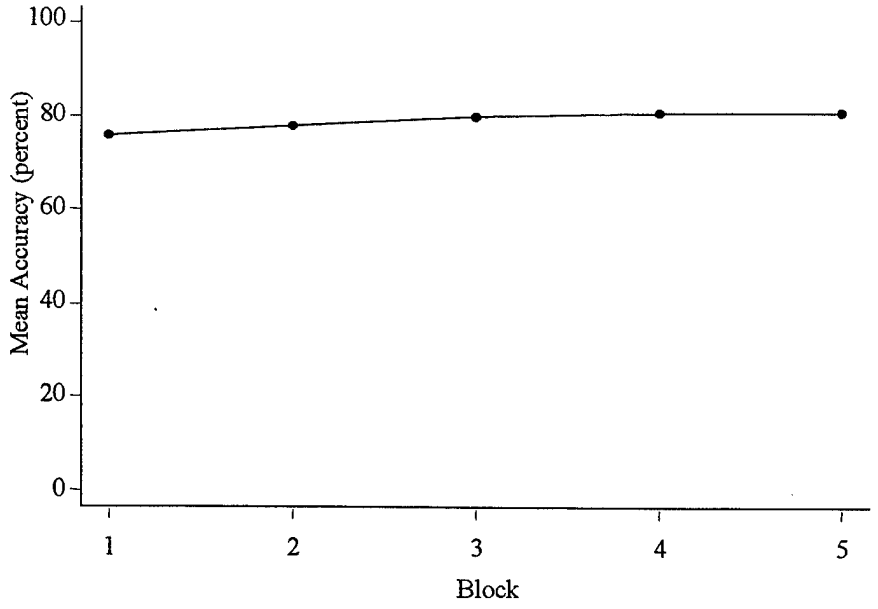


Figure 3.11. A significant effect for block (F (4,14665) = 18.8296, p = 0.0000) with accuracy as the dependent measure.

Significance could not be shown for any of the independent variables when analyzed separately: SOA by block (F (4,14665) = 1.8396, p = 0.1182), experience by block (F (4,14665) = 0.7282, p = 0.5725), scene category by block (F (16,14665) = 1.4198, p = 0.1217), orientation by block (F (4,14665) = 2.0148, p = 0.0895), sensor by block (F (1214665) = 1.2551, p = 0.2382). The failure of each independent variable to show statistical significance by itself further supports the conclusion that learning did not occur and helps validate the experimental design.

IV. CONCLUSIONS

The purpose of this thesis was to develop and execute a new methodology to investigate the human ability to detect objects imbedded in a visual scene. The approach broke down the search and detection process into fundamental components, the first of which is determining the effects of image fusion on global scene perception. The thesis investigated the effect of sensor, scene, experience, orientation and SOA on the ability to accurately determine the scene context.

The first major finding is that a significant sensor effect exists in determining scene context. The aggregate result shows that the IR scene provided a higher mean accuracy for overall scene comprehension than all other sensor formats, but fails to clearly establish the sensor format as consistently superior. The effect is biased by the sensor by scene interaction and supports previous target search and detection research (Buttrey et al. in preparation, Steele and Perconti 1997).

To understand the possible reasons that the IR component performs better than the other component or either fused image, a detailed look at the 23 individual scenes must be conducted. The review of the scenes shows that IR provides an accuracy advantage in eleven scenes, while fused color provides an accuracy advantage in ten scenes and I^2 provides an accuracy advantage in two scenes. In real world environments the reduction in workload and increased information available from having only one display coupled with the loss of target ambiguity will result in the fused color being more consistent than using two separate images.

The underlying factor affecting the ability of a sensor to outperform others is that the fused image performs as well as or better than either component image when, in addition to color, a greater number of "feature maps" are present in the scene. This finding supports the theory of Teichner et al (1979) and Wolfe (1994) that the presence of multiple features, or higher scene complexity, improves human visual preattentive processing. In this case, the more complex a scene, the higher the accuracy with which orientation can be preattentively determined.

Another major finding was that, based on statistical significance, the hypothesis that subject experience level has no effect on scene comprehension can be rejected. Practical evaluation of the results suggests that the means and variances for both novice and experienced subjects are the same. The actual values of mean accuracy vary less than one percent for IR and fused color and less than three percent for I^2 and fused monochrome. The fact that the sample distributions are similar in mean and variance raises doubts about rejecting the hypothesis that experience has no effect. If the practical aspect of significance is used, then the second hypothesis cannot rejected. This supports Biederman's (1983) theory that prior expectancy and familiarity do not show a benefit in scene perception.

The failure of experience to affect scene perception supports that the idea that the context of the scene is determined preattentively. Additionally, the length of the image presentation shows that the context of the scene is determined during the first visual fixation. Both of these show that a fused color image provides a subject with sufficient information to determine the scene context accurately during the first fixation. The subsequent failure of image fusion to consistently show a visual detection improvement (Krebs et al. 1997, Steele and Perconti 1997) suggests that this occurs during the guided search phase of visual detection.

The shortcoming of this study that should be addressed in future research is the limitations of the image data base. The images used were never intended for visual search experiments and do not fully encompass variations in real-world environments. As a result the images used are not composed of controlled and predetermined feature maps. If images containing specific features identified by Triesman (1985) and Wolfe (1994) can be accurately obtained, then fusion techniques to improve the enhancement of the features can be exploited. This thesis provides a unique methodology which can be used in continuing research on global scene perception.

APPENDIX A. TUKEYS PAIRED COMPARISONS

Sensor Combination	Estimate	Standard Error	Lower Bound	Upper Bound	Significance ($\alpha = 0.05$)
FC vs FM	0.0563	0.0112	0.0275	0.0851	yes
FC vs II	0.1150	0.0112	0.0863	0.1140	yes
FC vs IR	-0.0397	0.0112	-0.0685	-0.0109	yes
FM vs II	0.0588	0.0112	0.0300	0.0876	yes
FM vs IR	-0.0960	0.0112	-0.1250	-0.0672	yes
II vs IR	-0.1550	0.0112	-0.1840	-0.1260	yes

Tukey's method of multiple comparisons shows the mean accuracy for all two-way sensor comparisons is significant.

Scene Category Combination	Estimate	Standard Error	Lower Bound	Upper Bound	Significance ($\alpha = 0.05$)
Gen vs Man	-0.1810	0.01500	-0.2220	-0.1400	yes
Gen vs Road	-0.2600	0.01410	-0.0298	-0.2210	yes
Gen vs Sea	-0.0586	0.01500	-0.0995	-0.0180	yes
Gen vs Tree	-0.8190	0.01390	-0.1200	-0.0441	yes
Man vs Road	-0.0788	0.01120	-0.1090	-0.0483	yes
Man vs Sea	0.1220	0.01220	0.0889	0.1560	yes
Man vs Tree	0.0989	0.01080	0.0694	0.1290	yes
Road vs Sea	0.2010	0.01120	0.1710	0.2320	yes
Road vs Tree	0.1780	0.00962	0.1510	0.2040	yes
Sea vs Tree	-0.0233	0.01080	-0.0529	0.0063	no

Tukey's method of multiple comparisons shows the mean accuracy for two-way scene category comparisons is significant except for tree vs sea.

SOA Combination (milliseconds)	Estimate	Standard Error	Lower Bound	Upper Bound	Significance ($\alpha = 0.05$)
0 vs 20	0.0192	0.0125	-0.0534	0.0150	No
0 vs 40	0.0229	0.0125	-0.0570	0.1130	no
0 vs 60	-0.0346	0.0125	-0.0687	-0.0041	yes
0 vs 80	0.0459	0.0125	-0.0800	-0.0117	yes
20 vs 40	-0.0037	0.0125	-0.0378	0.0305	no
20 vs 60	-0.0154	0.0125	-0.0495	0.0188	no
20 vs 80	-0.0267	0.0125	-0.0608	0.0075	no
40 vs 60	-0.0117	0.0125	-0.0459	0.0224	no
40 vs 80	-0.0230	0.0125	-0.0572	0.0112	no
60 vs 80	-0.0113	0.0125	-0.0454	0.0229	no

Tukey's method of multiple comparisons shows the mean accuracy for two-way SOA comparisons of SOA 0 msec vs SOA 60 msec and SOA 0 msec vs SOA 80 msec to be significant. All other pairs cannot be rejected.

APPENDIX B. SUMMARY STATISTICS

Category	Factor	Mean Accuracy	Standard Error	N	Trial Images
Sensor	I^2	0.7522	0.2257	460	3680
	IR	0.8845	0.1410	460	3680
	FM	0.7899	0.1946	460	3680
	FC	0.8543	0.1587	460	3680
Experience Level	Experienced	0.8285	0.1840	920	7360
	Novice	0.8120	0.1957	920	7360
Orientation	Right Side Up	0.8065	0.1912	920	7360
	Upside Down	0.8340	0.1880	920	7360
Scene	Tree	0.7679	0.1956	560	4480
	Man Made	0.8668	0.1392	320	2560
	Road	0.9456	0.1022	480	3840
	Sea	0.745	0.1906	320	2560
	General	0.6859	0.2212	160	1280
SOA	0 msec	0.7979	0.2004	368	2944
	20 msec	0.8132	0.1998	368	2944
	40 msec	0.8183	0.1952	368	2944
	60 msec	0.8312	0.1838	368	2944
	80 msec	0.8407	0.1670	368	2944

APPENDIX C. SIGNIFICANT INTERACTIONS

	General	Man Made	Road	Sea	Tree
Right Side Up	67.656	86.563	94.688	74.609	72.411
Upside Down	69.531	86.797	94.427	74.297	81.161

The scene by orientation interaction means.

	Fused Color	Fused Monochrome	Image Intensified	Infrared
Right Side Up	83.696	79.293	74.619	85.000
Upside Down	87.174	78.696	75.815	91.902

The sensor by orientation interaction means.

	General	Man Made	Road	Sea	Tree
Fused Color	71.250	89.219	98.021	78.594	80.446
Fused Monochrome	70.313	84.688	94.167	69.688	70.536
Image Intensified	50.625	82.656	88.125	66.875	71.696
Infrared	82.188	90.156	97.917	82.656	84.464

The scene by sensor interaction means.

APPENDIX D. RANDOMIZED DESIGN

SOA (milliseconds)	Block one	Block two	Block three	Block four	Block five
0	6	6	7	6	7
20	7	7	6	6	6
40	6	7	6	7	6
60	6	6	6	7	7
80	7	6	7	6	6

The sequence of SOA's used in the experiment shown as number of times a specific SOA occurred in a given trial block over all sessions.

LIST OF REFERENCES

Biederman I., Teitelbaum R. C., Mezzanotte R. J., (1983). Scene perception: A failure to find a benefit from prior expectancy or familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 411-429.

Boyce S. J., Pollatsek A., (1992). Identification of objects in scenes: The role of scene background in object naming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 531-543.

Boyce S. J., Pollatsek A., (1992a). An exploration of the effects of scene context on object identification. In K. Rayner (Ed.), Eye Movements and Visual Cognition (pp. 227-242). New York, NY: Springer, Verlag.

Buttrey, S. E., Krebs, W. K., Lewis, P. A. W., McKenzie E., (in preperation). Pairwise Comparison of Natural Scenes. Unpublished manuscript. Naval Postgraduate School, Monterey, CA.

Coltheart M.,(1980). Iconic memory and visible persistence. Perception and Psychophysics, 27, 183-228.

Haber R. N. (1970). Visual information storage. Committee on Vision, Division of Behavioral Sciences National Research Council (pp. 129-150). National Academy of Sciences, Washington D. C.

Irwin D. E., Yoemans J. M., (1991). Duration of visible persistence in relation to stimulus complexity. Perception and Pychophysics, 50, 475-489.

Kahneman D., (1968). Method, findings and theory in studies of visual masking. Psychological Bulletin, 70, 404-425.

Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S., and Schuler, J. (1998). Beyond third generation: A sensor fusion targeting FLIR pod for the F/A-18. Proceedings of the International Society for Optical Engineering SPIE, Sensor Fusion: Architectures, Algorithms, and Applications II, Orlando, FL.

Neisser, U. (1967). Cognitive psychology. New York: Appleton, Century, Crofts.

Newman, E. A., Hartline, P. H. (1982). The infrared vision of snakes. Scientific American, 246, 116-127.

Palmer J., Ryan D., Tinkler R., Creswik H., (1993). Assessment of image fusion in a Night pilotage system. Multisensors and Sensor Fusion. Symposium conducted at North Atlantic Treaty Organization (NATO Ac/243 panel 3/4. Brussels Belgium.

Phillips W. A., (1974). On the distinction between sensory storage and short-term visual memory. Perception and Psychophysics, 16, 283-290.

Ryan D., Tinkler R., (1995). Night pilotage assessment of image fusion. In R. J. Lewandowski, W. Stephens, L. A. Haworth (Eds.), Proceedings of the International Society for Optical Engineering SPIE: SPIE-Vol. 2465. Helmet- and Head-Mounted Displays and Symbology Design Requirements II 1995. (pp. 50-67). Orlando Fl.

Scribner, D.A., Satyshur, M.P., and Kruer, M.R. (1993). Composite infrared color images and related processing. Proceedings of the IRIS Specialty Group on Targets, Backgrounds, and Discrimination.

Steele, P.M. and Perconti, P. (1997). Part task investigation of multispectral fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. Proceedings of the International Society for Optical Engineering SPIE: Vol. 3062 Aerospace/Defense Sensing, Simulation, and Controls (pp. 88-100). Orlando, FL.

Tactical jet night vision device manual (First ed.). (1994). Marine Aviation Weapons Squadron ONE, Yuma Az.

Teichner W. H., Mocharnuk J. B., (1979). Visual search for complex targets. Human Factors, 21, 259-275.

Therrien, C.W., Scrofani, J., and Krebs, W.K. (1997). An adaptive technique for the enhanced fusion of low-light visible with uncooled thermal infrared imagery. IEEE: International Conference on Imaging Processing, October 1997.

Toet A., IJspeert J. K., Waxmen A. M., Aguilar M. (1997). Fusion of visible and thermal imagery improves situational awareness. In J. G. Verly (Ed.), Proceedings of the International Society for Optical Engineering SPIE: Vol. 3088. Enhanced and Synthetic Vision 1997. (pp. 177-188). Orlando, FL.

Treisman A., (1985). Preattentive processing in vision. Computer Vision, Graphics, and Image Processing, 31, 156-177.

Treisman A., Sato S.,(1990). Conjunction search revisited. Journal of Expereimental Psychology: Human Perception and Performance, 16, 459-478.

Waxmen A. M., Gove A. N., Seibert M. C., Fay D. A., Carrick J. E., Racamato J. P., Savoye E. D., Burke B. E., Reich R. K., McGonagle W. H. & Craig D. M.,(1996). Progress on color night vision: Visible / IR fusion, perception and search, and low-light CCD imaging. In J. G. Verly (Ed), Proceedings of the International Society for Optical Engineering SPIE: Vol. 2736. Enhanced and Synthetic Vision 1996. (pp. 96-107). Orlando Fl.

Waxmen A. M., Gove A. N., Fay D. A. , Racamato J. P., Carrick J. E. , Seibert M. C., Savoye E. D.,(1996a). Color night vision: Opponent processing in the fusion of visible and IR imagery, Neural Networks 10, 1-6.

Waxmen A. M., Fay D. A., Gove A. N., Seibert M. C., Racamato J. P., Carrick J. E. & Savoye E. D., (1995). Color night vision: Fusion of intensified visible and thermal IR imagery. In J. G. Verly (Ed.), Proceedings of the International Society for Optical Engineering SPIE: Vol. 2463. Enhanced and Synthetic Vision 1995. (pp. 58-68). Orlando Fl.

Wolfe J. M., (1994). Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin, 2, 202-238.

Wolfe J. M., (1993). Visual search in continuous, naturalistic stimuli. Vision Research, 34, 1187-1195.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Rd. Ste 0944
   Fort Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey, CA 93943-5101

3. Professor William K. Krebs 4
   Code OR/Kw
   Naval Postgraduate School
   Monterey, CA 93943-5002

4. Professor Harold J. Larson 1
   Code OR/La
   Naval Postgraduate School
   Monterey, CA 93943-5002

5. Lieutenant Commander Brice L. White 2
   1342 W. Princess Anne Road
   Norfolk, VA 23507

6. Naval Research Laboratory 1
   Attn: D. Scribner, Code 5636
   Washington, DC 20378

7. Commanding Officer 1
   MAWTS-1
   Box 99200
   Yuma AZ 85369-9200

8. Office of Naval Research 1
   Attn: Dr. Joel Davis
   800 North Quincy Street
   Arlington, VA 22217-5660

9. Lockheed Martin E&M 1
   Attn: Mr. Robert McDaniel
   5600 Sand Lake Road, MP 126
   Orlando, FL 32819-8907